Planning, Practice & Research, Vol. 18, No. 2–3, pp. 213–217, May–August, 2003

Carfax Publishing
Taylor & Francis Group

PRACTICE FORUM

GIS, the US Census and Neighbourhood Scale Analysis

MARC SCHLOSSBERG

Introduction

One current focus of planning analysis is at the neighbourhood level—understanding local area phenomena and developing plans and policies to increase the quality of life at the very local scale. Geographic Information Science (GIS) adds a layer of capacity to such analysis by making rich disaggregated datasets accessible in a way (through maps) not possible previously. As these tools are used for small area, neighbourhood-oriented analysis and decision making, it is important for planners to understand how the choices one makes in the most fundamental parts of GIS-based neighbourhood analysis (data scale and spatial analysis technique) affect subsequent results.

Data from the 2000 US Census is easily accessible to anyone with a computer and an Internet connection. Likewise, the census spatial data (the census outlines or polygons) are also freely available to download in GIS-compatible format. Thus, acquiring both the census outlines (tracts, block groups, blocks) and the attribute data (population, poverty, housing units, etc.) that gives meaning to the census polygons is no longer an issue in conducting small area socio-demographic analysis. However, how do you choose what scale of census data to include and which GIS tool to use in the spatial analysis? Or more fundamentally, to what degree does the type of spatial analysis method and scale of census unit matter when conducting GIS-based small area demographic analysis? Two brief exercises presented below are designed to provide insight into issues of scale and basic spatial analysis techniques that are commonly employed, yet often misused by planning students and practicing professionals. Using these basic tools and datasets incorrectly, in calculating the number of people in close proximity to toxic sites for example, can have significant impacts on policy decisions that emerge partially from spatial analyses on population and housing data.

The utilisation and analysis of census or other small area data must be performed appropriately or the analysis may be skewed or misleading. Recently, there have been warnings of the dangers of inappropriate representation of data in map form (Kent & Klosterman, 2000) and new ways to teach planning-relevant GIS have been presented (Montagu, 2001), but an equal warning about appropriate utilisation of census data and spatial extraction techniques has not been recently made. How the choice of different combinations of data and analysis technique can create significantly different results is not a new concept. The modifiable areal unit problem (MAUP), as this phenomena is called, was introduced in 1979 (Openshaw & Taylor, 1979) and the concept dates back to 1934 (Gehlke & Biehl, 1934). The MAUP is "a problem arising from the imposition of artificial units of spatial reporting on continuous geographical phenomenon resulting in the generation of artificial spatial patterns" (Heywood *et al.*, 1998). Despite this history of knowledge about the topic, it is rare to find such understanding present in current planning practice and education. Thus, it is critical that such knowledge continue to be actively articulated as GIS and census analysis become more ac-

Marc Schlossberg, Planning, Public Policy, and Management Department, University of Oregon, 128 Hendricks Hall, Eugene, OR 97403, USA. Email: schlossb@uoregon.edu

0269-7459 Print/1360-0583 On-line/03/020213-05 © 2003 Taylor & Francis Ltd
DOI: 10.1080/0269745032000168269

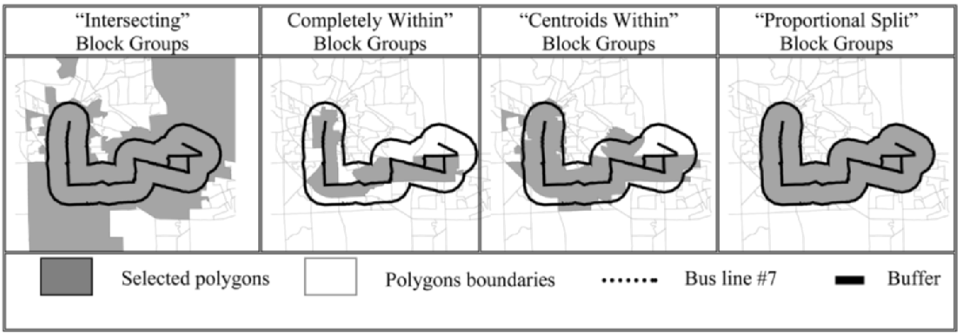

FIGURE 1. Comparison of spatial extraction results.

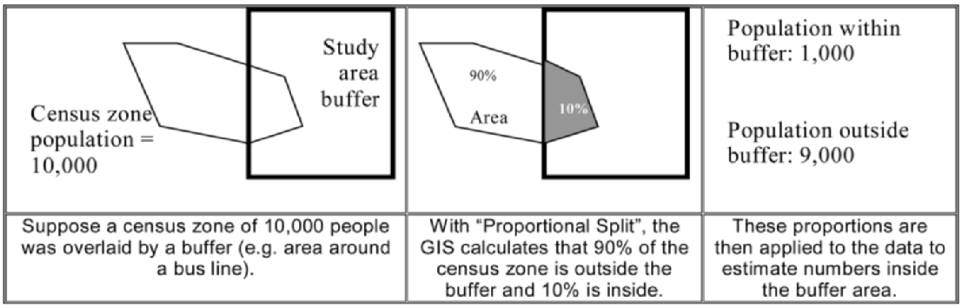

FIGURE 2. Explanation of 'proportional split'.

cessible to users without much formal Geographic Information Science training.

Two fairly simple and common types of GIS-based census analyses were conducted to illustrate the importance of understanding census scale and spatial extraction methodology. Census data at three scales (blocks, block groups and census tracts) were acquired and then analysed utilising four methods (completely within, centroids within, intersecting, proportional split) in combination with an overlay zone of a transit line in one example and a neighbourhood boundary in another. Census tracts, block groups, and blocks are all three different spatial categorisations by which census (or enumeration area) data are aggregated in the USA. On average, a block contains 65 people, a block group contains 1100 people, and a census tract contains 4000 people. In terms of the spatial extraction techniques, the 'completely within' method chooses only those census zones that are entirely within the study area; the 'centroids within' method includes those census zones that have their geographic

TABLE 1. Population results for transit zone scenario

	Intersecting	Completely within	Centroids within	Proportional split	Variation
Census tracts	116 432	15 343	50 952	54 147	*+/- 101 000*
Block groups	94 581	32 500	55 664	54 781	*+/- 62 000*
Census blocks	55 856	39 295	45 828	56 139	*+/- 16 000*
Variation	*+/- 61 000*	*+/- 23 000*	*+/- 10 000*	*+/- 2000*	

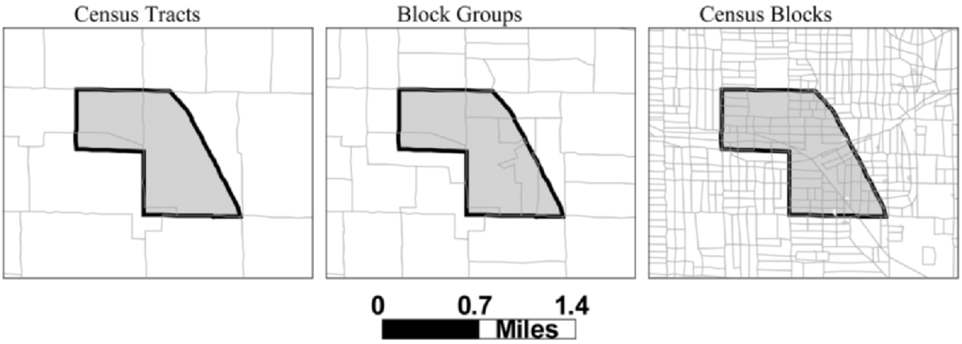

FIGURE 3. Oakdale neighbourhood overlaid with different census units.

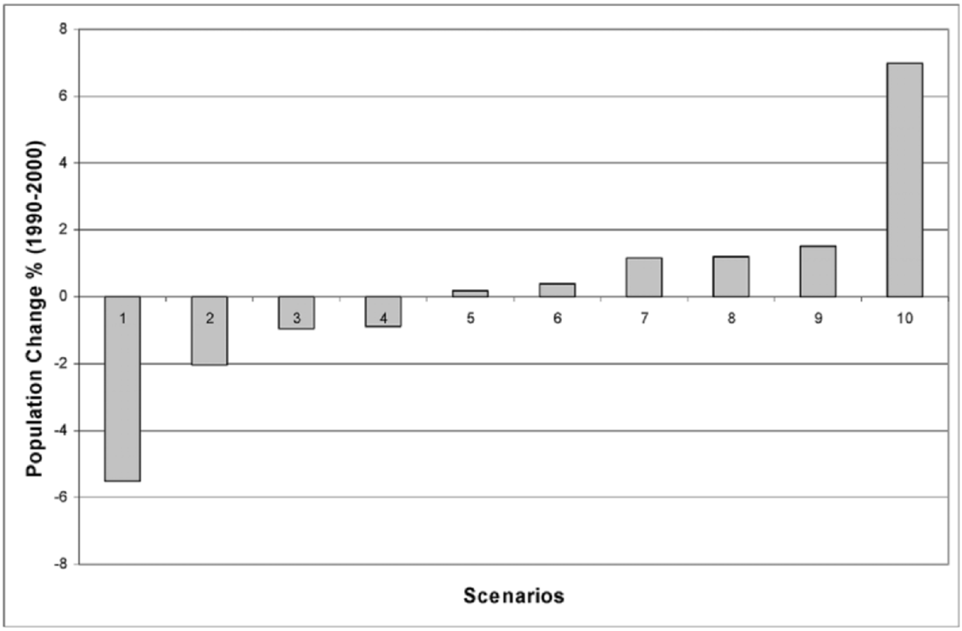

FIGURE 4. Population change in Oakdale neighbourhood (1990–2000). All census units and extraction methods sorted from lowest to highest change.

centre within the study area; the 'intersecting' method selects all census zones that touch the study area; and the 'proportional split' method applies the proportion of a census zone that is within the study to the census data of that particular census zone. Figure 1 shows a transit line buffer zone ('study area') with the census zone selected using each of the four different types of spatial extraction methods.

The images in Figure 1 illustrate that the set of results obtained using the different spatial extraction methods varies considerably. In the 'proportional split' method, the selected census units exactly conform to the shape of the buffer, although this is somewhat misleading in terms of the results this method produces. Figure 2 illustrates how this method creates estimates of the underlying population and housing data based on the geographic proportion of a census zone that falls inside of the buffer area.

Using these four spatial extraction methods and applying them to three different census scales (census tract, block group and block), 12 results were produced, one for each research

TABLE 2. Pros and cons of data and extraction techniques

Census Units	Pro	Con
Blocks	Small size; good for neighbourhood level analysis	Limited variables available; depending on size of overall geographic study area, extra computing resources may be needed
Block groups	Relatively small in size; can be used for large neighbourhoods or sub-city level analysis; all census variables available	Too large for many neighbourhood level analyses, but tempting to use because of large availability of data
Census tracts	All census variables available; fewer computing resources necessary; good for sub-county analysis	Too big for neighbourhood or small area analysis
Spatial Extraction Method		
Intersecting	Exhaustive: includes all polygons that fall within the study area	Too inclusive
Completely within	Good targeting: no polygons outside of the study area will be included	Too exclusive
Centroids within	Good compromise: polygons with a majority of its area within the study area will be included	Calculating centroids is imprecise and adds additional error
Proportional split	Forces selected polygons to mirror the shape of the study area	Assumes uniform dispersion of attributes within each polygon

scenario. Table 1 lists the population living in the transit area study zone, using the different types of census units and different methods of spatial analysis. The figures clearly show that in this one example, the population living within the transit zone ranges from 15 000 to 116 000 the population living within the transit zone ranges from 15 000 to 116 000 people (!), depending on the combination of census scale and extraction method used.

As another experiment, an analysis of population change from 1990 to 2000 was conducted using a pre-defined neighbourhood boundary (Oakdale in Grand Rapids, MI), see Figure 3.

Applying the 12 analyses to this neighbourhood example and ranking the results from lowest to highest in terms of population change, Figure 4 shows that depending on the combination of census unit and extraction technique, the neighbourhood could be characterised as either gaining or losing population over the last decade. Note that two of the scenarios are absent because there are no census tracts or block groups entirely within the neighbourhood boundary (using such a method would have the result of showing zero people living in this neighbourhood).

A follow-up set of calculations was conducted on each experiment using the total area of the selected census units and comparing it to the total area of the study area zone to see which combination of census unit and spatial extraction method produces results closest to the 'true' study area size. Results from these two experiments (transit buffer and neighbourhood analysis) and the two sets of calculations (population and total area) lead to the following conclusions (see Table 2).

The basic conclusion from the above analysis is that, when possible, use the smallest areal unit available (a finding consistent with the Modifiable Areal Unit Problem or MAUP). While such a conclusion may seem obvious, census tracts seem to remain the scale of choice by many practitioners and researchers not trained in the Geographic Information Sciences. In this case, census blocks are the smallest unit. Using census blocks can be limiting, however, because there are hundreds of additional variables available at the block group and census tract level that are not available at the block level. So if the additional variables only available at these larger administrative scales are of interest, then block groups should be used instead of census tracts, especially in sub-county analysis. There is very little to be gained by using census tracts and much to be lost in terms of reliability and accuracy of the data when used in combination with one of the spatial analysis and extraction techniques discussed here.

The other conclusion to be gleaned from this analysis is that when available, the 'proportional split' method of extracting data under a buffer, neighbourhood boundary or similar administrative overlay zone is the most reliable. That is, the results produced by this method show the greatest consistency. The proportional split method is not always an available option within GIS software (or is not as obvious as with the other techniques). Therefore, when this option is not available or easily accessible to the beginning or intermediate user (i.e. most planners and planning students), then selecting polygons that have their centroid, or geographic centre, within the study buffer or zone should be used.

References

Gehlke, C. E. & Biehl, K. (1934) Certain effects of grouping upon the size of the correlation coefficient in census tract material, *Journal of the American Statistical Association, Supplement*, 29, pp. 169–170.

Heywood, D. I., Cornelius, S. & Carver, S. (1998) *An Introduction to Geographical Information Systems* (New York, Addison Wesley Longman).

Kent, R. B. & Klosterman, R. E. (2000) GIS and mapping: pitfalls for planners, *Journal of the American Planning Association*, 66(2), pp. 189–198.

Montagu, S. A. (2001) Repackaging the revolution: making GIS instruction relevant to planners, *Journal of Planning Education and Research*, 21(2), pp. 184–195.

Openshaw, S. & Taylor, P. (1979) A million or so correlation coefficients: three experiments on the modifiable areal unit problem, in: N. Wrigley (Ed.) *Statistical Applications in the Spatial Sciences*, pp. 127–144 (London, Pion).